

РЕАЛИЗАЦИЯ КОНТЕКСТНЫХ СООТВЕТСТВИЙ *Ы:ы, Ы:е* и *Ы:0* В ФАЙЛЕ ФОНОЛОГИЧЕСКИХ ПРАВИЛ

Р.А.Гильмуллин

Казанский государственный университет
Rinat.Gilmullin@ksu.ru

Аннотация. В этой статье описываются фонологические правила представления соответствия лексического символа *Ы* татарского языка поверхностным буквам *ы, е* и *0* в словоформе в зависимости от контекста, т.е. от предыдущих и последующих символов. Как уже отмечалось в первой статье, файл фонологических правил является частью двухуровневого морфологического анализатора, который используется как на этапе генерации словоформы из составляющих ее морфем, так и на этапе распознавания — разложения словоформы на корректную последовательность составляющих ее морфем на основе файла морфотактических правил. Программа генерации словоформ на основе файла правил реализована как автономный модуль и может быть использована в программах перевода, человеко-машинного интерфейса и морфологической коррекции текстов, а также, как подсистема обучения морфологии татарского языка.

Введение. Данная статья является второй из серии запланированных, в которых будут детально описываться правила генерации словоформ из составляющих их морфем, и посвящена раскрытию трех последующих правил относительно соответствий *Ы:ы, Ы:е* и *Ы:0*, которые также, как и первые три правила соответствия, в основе своей, отражают закон сингармонизма.

Интерпретация фонологических правил татарского языка на примере 3-х правил, описывающих соответствие *Ы:ы, Ы:е* и *Ы:0*

Также, как и для лексического символа *А*, блок правил соответствия лексического символа *Ы*, в основе своей, отражает закон сингармонизма. Однако в ряде случаях, такая замена вызвана нарушениями в литературном языке из-за кириллической графики, или нарушения правил ассимилирования в заимствованных словах. Последнее соответствие *Ы:0* означает, что лексический символ *Ы* в поверхностной форме аннулируется, т.е. просто опускается. В

таблице 2 приведен фрагмент описания соответствий морфем с лексическим вхождением символа **Ы** и их алломорфов. Как отмечалось в первой статье, некоторые морфемы могут включать несколько лексических символов, для каждого из которых имеется свое фонологическое правило представления соответствий.

Таблица 2. Фрагмент описания соответствия морфем и их алломорфов.

Лексическая запись	Поверхностная запись
-Ыш	-ыш, -еш
-Ым	-ым, ем, -м
-Ыштыр	-ыштыр, -ештер, -штыр, -штер
-ДЫр	-дыр, -дер
-Ырга	-ырга, ергә, -рга, -ргә
-мЫн	-мын, -мен
-ДЫ	-ды, -де
-Ыр	-ыр, -ер, -р
-СЫ	-сы, -се, -ы, -е
-нЫ	-ны, -не, -н
...	...

Правило (4) для лексического символа **Ы** имеет вид:

(4) ПРАВИЛО Ы:ы => VOWEL:BACKV CS:@ * @:0 * +:0 * _ | VOWEL:BACKV CS @:0 * +:0* CS* _|@:0* W:й +:0 * CS:@* _ | W:a CS:@* +:0 CS:@* _ | W:a +:0 Д:д _ | W:a (+:0) CS:@_ | е `:0 CS:@ * +:0 @:0* CS* _ | [#|-] д у с [+т|+:0] C:0* _|[#|-] а с [+т|+:0] C:0* _;

Символ | обозначает логическое **ИЛИ**. Следовательно, **ПРАВИЛО (4)** включает 8 левосторонних контекстов, где возможно соответствие **Ы:ы**:

- VOWEL:BACKV CS:@ * @:0 * +:0 * _;
- VOWEL:BACKV CS @:0 * +:0* CS* _;
- @:0* W:й +:0 * CS:@* _;
- W:a CS:@* +:0 CS:@* _;
- W:a +:0 Д:д _;
- W:a (+:0) CS:@_;
- е `:0 CS:@ * +:0 @:0* CS* _;
- [#|-] д у с [+т|+:0] C:0* _|[#|-] а с [+т|+:0] C:0* _;

Рассмотрим **Правило (4)** соответствия **Ы:ы** в контексте (а). Согласно структуры фонологического правила оно утверждает следующее: лексический символ **Ы** соответствует поверхностному символу **ы**, если ему предшествует любая гласная из множества заднего ряда гласных **BACKV**; далее любой символ из множества **CS**; затем случай, когда при присоединении какого-либо аффикса в татарском языке происходит выпадение символа или аннулирование на поверхностном уровне; далее может следовать символ **+** для присоединения аффиксов.

Приведем примеры с контекстом (а). Пусть на вход генератора подаются следующие последовательности морфем:

ат+Ыш 'стреляй' + [аффикс взаимного залога]

ат+Ым 'лошадь' + [аффикс притяжательности]

В процессе генерации по файлу фонологических правил устанавливаются промежуточные соответствия **а:а, т:т, +:0**, для которых нет специальных правил соответствия; далее работает **Правило (4)**, левый контекст для которого удовлетворяет контексту (а) и, следовательно, **Правило (4)** устанавливает соответствие **Ы:ы**.

Лексический уровень (Л.у.): *а т + Ы ш* *а т + Ы м*

Поверхностный уровень (П.у.) *а т 0 ы ш* *а т 0 ы м*

Генератор нам выдаст следующие поверхностные формы записи:

атыш 'стрельба, помощи стрелять'

атым 'моя лошадь'

Далее рассмотрим контекст (б) **Правила (4)**:

VOWEL:BACKV CS @:0 * +:0* CS*_

Контекст (б) включает в себе описание структуры множества слов, которые в основе своей имеют символы из множества заднего ряда гласных и заканчиваются символами из множества **CS**, затем следует любой символ, который может аннулировать. Далее следует символ присоединения аффикса, после которого - символ из множества **CS**.

Главным образом, это контекст для группы аффиксов, где лексическому символу **Ы** предшествует символ из множества **CS**.

Примеры:

кош+СЫ+нЫ 'птица' + [афф. прит. 2 л.] + [аккузативный надеж]

бар+ДЫ 'иди' + [афф. прошедшего категорического времени]

Оба примера лексического соответствия **Ы:ы** удовлетворяют контексту (б) **Правила (4)**. Генератор выдаст следующие поверхностные формы записи:

кошын 'его птицу'

барды 'сходил'

Далее соответствие **Ы:ы** анализируется в 4 контекстах с лексического символом **W**, который является аффиксом, образующим форму глаголов настоящего времени, и имеет свое фонологическое правило соответствия.

(в) @:0* W:й +:0 * CS:@*_;

Согласно контексту (в) соответствие **Ы:ы** проявляется тогда, когда ему слева может предшествовать любой символ из множества **CS**, далее символ присоединения аффиксов +, возможно, переходящий в 0, затем лексический символ **W**, который на поверхностном уровне соответствует символу **й**. Далее следует любой символ, который может аннулироваться. Ниже в примере - это символ **а**.

Пример:

кара+W+cЫң 'смотри' + [афф. наст. времени] + [афф. 2-го лица ед. числа]

Поверхностная форма будет иметь вид:

карыйсың 'смотришь'

Рассмотрим далее три правила, связанные с соответствием **W:a**.

г) W:a CS:@*_ +:0 CS:@*_;

д) W:a +:0 Д:д_;

е) W:a (+:0) CS:@*_;

Контекст (г) означает, что слева от лексического символа **Ы** может быть любой символ из множества **CS**, который соответствует любому символу из алфавита, далее уже известный нам символ +. Символу + снова может предшествовать любой символ алфавита, далее - лексический символ **W**.

Пример:

кайт+W+bЫз 'возвращайся' + [афф. наст. времени] + [афф. 1-го лица мн. числа]

Левый контекст для соответствия **Ы** удовлетворяет условию контекста (г) и на выходе генератора получим:

кайтабыз 'возвращаемся'

Контексты (д) и (е) являются частным случаем контекста (г), когда предшествующие символы нельзя заранее определить в списке вспомогательных множеств. Для контекста (д) это лексический символ **Д**, а контекст (е) обеспечивает реализацию лексического соответствия **Ы:ы**, когда лексические символы **W** и **Ы** связаны в одной морфеме -**WcЫ**.

Пример:

бар+W+ДЫр 'иди' + [афф. наст. времени] + [афф. модального вопроса]

бор+WcЫI ' поверни ' + [афф. условного будущего времени]

Благодаря контекстам (д) и (ё) генератор выдаст следующие поверхностные формы:

барадыр ' он, наверное, пойдет? '

борасы (бар) ' (надо) будет повернуть '

Рассмотрим контекст (ж) и Правило (7):

ж) е':0 CS:@ * +:0 @:0* CS*_;

(7) ПРАВИЛО Ы:е /<=':0 CS:@ * +:0 @:0* CS*_;

Также как и в первой статье, где описывалась реализация контекстных соответствий *A:a* и *A:э*, этот контекст описывает исключительные ситуации, возникающие, в основном, из-за не ассимилированных заимствований с русского, английского и других языков, или иных нарушений закономерностей татарского языка [4]. Таким образом, данный контекст описывает лексическое соответствие, нарушающее закон сингармонизма. Слова, которые изменяются с нарушением сингармонизма записываются в корневом словаре морфологического анализатора с дополнительным символом ' (апостроф).

Пусть на вход генератора поступает:

бале't+ЫIм ' балет ' + [аффикс притяжательности]

Такой контекст будет удовлетворять условию контекста (ж) и генератор выдаст:

балетым ' Мой балет '

На самом деле генератор выдал бы и вторую неправильную поверхностную форму *балетем*, т.к. такой контекст, последний гласный символ которого принадлежит множеству **FRONTV** переднего ряда гласных, удовлетворяет аналогичному контексту (а) **Правила (4) в Правиле (5)**. Для предотвращения этого вводится дополнительное **Правило (7)**, которое утверждает, что соответствие **Ы:е** никогда не проявляется в данном контексте. Рассмотрим контекст (з):

[#|-] д у с [+;т|+:0] C:0*_|[#|-] а с [+;т|+:0] C:0*_;

Для некоторых слов, при употреблении их с аффиксом притяжательности 3 лица, ед. числа, наблюдается добавление поверхностного символа *т*. Согласно данного контекста (з), в на поверхностном уровне символ + может быть заменен этим символом.

Пример:

дус+СЫI ' друг ' + [аффикс притяжательности]

ас+СЫI ' низ ' + [аффикс притяжательности]

Результатом генератора будут следующие поверхностные формы:

дусты 'его друг'
асты 'то, что является низом'

Правило (5) содержит множество контекстов, где выполняется соответствие **Ы:е**.

(5) **ПРАВИЛО Ы:е** => [VOWEL:FRONTV | %:a|Q:I|(+:0) CS:@ * @:0 * +:0 * _ | VOWEL:FRONTV CS(CS) @:0 * +:0 * CS* _ |@:0* W:i +:0* CS:@* _ | W:A (+:0) CS:@ _ |W:A +:0 D:d _ |[#|-] ɐ s +:t C:0* _ ;

Здесь имеется 5 контекстов соответствия **Ы:е**.

а) [VOWEL:FRONTV | %:a|Q:I|(+:0) CS:@ * @:0 * +:0 * _ | ;

б) VOWEL:FRONTV CS(CS) @:0 * +:0 * CS* _ | ;

в) @:0* W:i +:0* CS:@* _ | ;

г) W:A (+:0) CS:@ _ | ;

д) W:A +:0 D:d _ | ;

е) [#|-] ɐ s +:t C:0* _ ;

Правило (5) определяется также как и **Правило (4)**, с той лишь разницей, что вместо множества заднего ряда гласных **BACKV** используется множество переднего ряда гласных **FRONTV**. Для разрешения особых случаев, вместо символа ` (апостроф), как это принято в контексте (ж) **Правила (4)**, в первом контексте **Правила (5)** используется символ **Q**. Кроме того, используется символ '%', который тоже вводится из-за нарушения закона сингармонизма.

Примеры.

Пусть на вход генератора подаются следующие лексические формы:

бэл%+ЛАр+Ым 'беда' + [афф. множ.] + [афф. притяж.]

кил+ДЫ 'приходи' + [афф. прошедшего категорического времени]

сөйлә+W+сЫң 'рассказывай' + [афф. наст. времени] + [афф.

2-го лица ед. числа]

кит+W+ДЫр 'уходи' + [афф. наст. времени] + [афф.

модального вопроса]

бир+W+сЫ (бар) 'дай' + [афф. условного будущего времени]

әс+сЫ 'верх' + [аффикс притяжательности]

шигQрь+Ым 'стих' + [аффикс притяжательности]

Генератор, соответственно, выдаст следующие поверхностные формы:

бәләләрем 'мои беды'

килде 'пришел'

сөйләсең 'рассказываешь'

китәдер 'наверное, уходит'

бирәсе (бар) '(надо) будет отдать'

‘мой стих’

Рассмотрим **Правило (6)** лексического соответствия **Ы:0**.

$$+ : 0 \text{ [C:s|C:0] [BI:I|BI:e] } + : 0 \text{ n } ;$$

Оно состоит из 2 контекстов:

(a) `[@:a|e:e|l:l|o:o|ø:ø|Ы:ы|Ы:e|@:A|%:a|+|_]|0`;

(6) $+ : 0 [C:c|C:0] [bI:I|bI:e] + : 0 n$;

Контекст (а) утверждает, что лексический символ **Ы** соответствует

Пример:

[афф. притяж. 1л. ед.ч.]

Первый лексический символ **Ы** в данной словоформе встречается в

илтәсем (бар) 'что-то нужно будет отнести'

H;

Запись означает, что соответствие **Ы:0** проявляется, если слева от

Пример:
 эти+СЫ+НЫ 'отец' + [афф. притяж. 3л. ед.ч] + [афф. аккузат. надежа]
 ул+СЫ+НЫ 'сын' + [афф. притяж. 3л. ед.ч] + [афф. аккузат. надеж]

Для обоих примеров первое соответствие **Ы:е** и **Ы:ы** обеспечиваются вторыми контекстами **Правил (5)** и **(4)**, соответственно. Второе соответствие лексического символа **Ы** поверхностному символу **0** происходит благодаря контексту **(6)** **Правил (6)**, следовательно, генератор выдаст следующие поверхностные формы записи:

этисен	'его отца'
улыи	'его сына'

Заключение. Файл фонологических правил татарского двухуровневого морфологического анализатора составлен из 42 правил, 11 из которых реализуют ситуации, вызванные нарушениями в татарской морфологии. Полный список фонологических правил татарского языка приводится в работе [1].

В статье подробно рассмотрены 3 правила, а именно - правила реализации соответствия лексического символа **Ы** поверхностным символам **ы**, **е** и **0**. Следующая статья будет посвящена описанию фонологических правил представления лексического символа **У** поверхностным символам **у**, **у** и **в**, лексического символа **Д** поверхностным символам **и**, **д** и **т**, лексического символа **Л** поверхностному символу **и** и лексического символа **С** поверхностному символу **с**.

ЛИТЕРАТУРА

1. Suleymanov D.Sh., Gilmullin R.A., Guilmy A.A. *A Two-level phonological rules of tatar morphology* KDS-97. Шестая Международная конференция «ЗНАНИЯ - ДИАЛОГ - РЕШЕНИЕ». Сборник научных трудов в двух томах. Ялта - 1997. Том 1. -с.299-305.
2. Evan L. Antworth. «PC-KIMMO: A Two-level Processor for Morphological Analysis SUMMER INSTITUTE OF LINGUISTICS. Occasional Publications in Academic Computing, 1990.
3. *Татарская грамматика* под ред. Закиева М.З.Том 2. Морфология.
4. Сулейманов Д.Ш. *Регулярность морфологии татарского языка и типы нарушений в языке*. -Казань. Изд-во КГУ. В сб. Трудов «Интеллект. Язык. Компьютер», вып.1,1994. -с. 77-106.